

USING SOUND TO UNDERSTAND BIOLOGICAL SEQUENCE DATA

Edward J. Martin¹, Daniel Barker¹, and Thomas Robert Meagher²

¹University of Edinburgh, UK

²University of St Andrews, UK

e.j.martin@sms.ed.ac.uk, Daniel.Barker@ed.ac.uk

trm3@st-andrews.ac.uk

ABSTRACT

Advancing DNA sequencing technology has created vast, ballooning databases of DNA/protein sequence data. DNA/proteins are polymers, composed of monomers (bases/amino acids respectively). Each monomer has different properties, and the sequence of monomers is responsible for the biological role of the polymer. Letters are used to represent the monomers, and thus DNA/protein sequences are represented as words.

Automated analyses of polymers are proliferating in bioscience — however knowledge discovery by these methods requires researcher corroboration. Visualisation is a key tool for representing sequences. Protein sequence visualisation typically relies on colour schemes to represent the complex properties of different amino acids. Though valuable, visualisation software can be impractical, particularly with datasets growing.

Sonification has potential benefits as an alternative or complement to visualisation. The linear, discrete nature of DNA/protein sequence data is highly amenable to parameter-mapping Sonification.

We have developed methods for DNA/protein sonification in single, paired, and grouped contexts utilising algorithmic mappings from data to sound. We also produce artistic compositions inspired by these mappings. We investigate these techniques for knowledge discovery and science communication. We believe leveraging psychoacoustic intuition and enthusiasm for sonification can improve bioinformatics. Our talk will discuss our current results and approaches.

<https://sonified.com/bmc-bioinformatics-2021>

<https://vimeo.com/521340925>

<https://vimeo.com/470163206>

<https://vimeo.com/485902033>

<https://vimeo.com/509161022>

